# The Seed Protein, Oil, and Yield QTL on Soybean Linkage Group I

J. Chung, H. L. Babka, G. L. Graef, P. E. Staswick, D. J. Lee, P. B. Cregan, R. C. Shoemaker, and J. E. Specht*

## ABSTRACT

**Soybean [*Glycine max* (L.) Merr.] seed protein is negatively correlated with seed oil and often with yield. Our goal was to examine the basis for these correlations at a quantitative genetic locus (QTL) level. Seventy-six $F_5$–derived recombinant inbred lines (RILs) from the mating of the high-protein (480 g protein per kg seed) *G. max* accession PI 437088A with the high-yield cultivar Asgrow A3733 (420 g kg$^{-1}$ seed protein content) were evaluated in six irrigation treatments (i.e., 100, 80, 60, 40, 20, and 0% replacement of weekly evapotranspiration loss) of a two-replicate experiment conducted for 2 yr. The RILs were genotyped with 329 random amplified polymorphic DNAs (RAPDs), 103 simple sequence repeats (SSRs), and four other markers, creating a 2943-centimorgan (cM) genetic map of 35 linkage group (LGs) that, on the basis of SSR homology, aligned with the 20 known soybean LGs. The phenotypic regression of RIL protein and oil on yield revealed respective linear coefficients of −2.6 and +1.6 percentage points per kg ha$^{-1}$ (a protein/oil exchange ratio of −1.6). A seed protein, oil, and yield QTL mapped close to RAPD marker OPAW13a in a small LG-I interval that was flanked by the SSR markers Satt496 and Satt239. The additive effects of the PI 437088A allele on seed protein, oil, and yield were a respective +1.0 and −0.6 percentage points (a protein/oil ratio of −1.6) and −154 kg ha$^{-1}$. Given that the genetic-based protein/oil exchange ratio of 1.6 is smaller than the 2.0 calorific-based oil/protein ratio, one might expect the remaining 0.4 units of carbon and/or energy to be made available for other seed dry matter. However, yield almost invariably falls when seed protein is genetically enhanced at the expense of seed oil, suggesting that protein synthesis and its deposition in the seed is energetically more costly to the plant than is commonly assumed.**

The predominant source of seed protein in the USA is the soybean. Before the late 1950s, oil was the seed constituent of economic importance. However, markets rapidly developed for the protein meal byproduct of soybean oil processing. By the early 1970s, 60% of the value of the unprocessed soybean seed was derived from the protein meal (Hartwig, 1973). During the past 16 yr (1986–2001), the U.S. soybean crop has averaged, on a 130 g kg$^{-1}$ seed moisture basis, a seed yield of 2401 kg ha$^{-1}$, a seed protein content of 354.1 g kg$^{-1}$, and a seed oil content of 185.6g kg$^{-1}$ (Hurburgh, 2001). These statistics indicated that 67.25 kg of the crop could be processed into 19.7 kg of 48% protein meal product and 4.90 kg of oil product.

Soybean protein and oil contents in various regions of the USA can deviate significantly from the foregoing national averages. Some of that geographic variability arises from meteorological events that, in any given season, randomly affect some regions but not others. For example, high temperatures during soybean seed development can elevate seed oil (Howell and Cartter, 1958), while severe drought can depress seed protein (Specht et al., 2001). Other geographic variability arises from region-specific climatic parameters. Notably, seed protein is typically lower in the northwestern than in southeastern soybean-growing states. In 2001, that regional difference spanned 3 percentage points, the largest ever observed in 17 yr of survey data (Hurburgh, 2001). Processors may not be able to derive the valuable 48% soybean meal if the soybean seed has too low of a protein content. To offset this geographical disadvantage, breeders developing high yielding cultivars adapted to northern and western production regions must practice coordinate selection for intrinsically greater seed protein.

Among the accessions in the USDA soybean germplasm collection (NGRP, 2001), seed protein ranges from 347 to 552 g kg$^{-1}$ on a dry seed basis ($\mu$ = 435 g kg$^{-1}$; $n$ = 11779), while seed oil ranges from 65 to 287 g kg$^{-1}$ ($\mu$ = 184 g kg$^{-1}$; $n$ = 11775). Such wide ranges would suggest that there is adequate genetic variability for breeders to develop cultivars high in *both* protein and oil. However, only 1750 (about 15%) of the 11 775 germplasm accessions have a protein *and* an oil content at or above their respective population means. Just 78 (0.66%) qualify if the criterion is ≥1.05× both population means, and only one accession qualifies at a ≥1.10× criterion. Even among the multientry trials conducted each year in soybean-producing states, it is rare to find any entry whose seed protein and oil content exceeds both respective constituent trial means (Thompson et al., 2001). Highly negative phenotypic and genotypic

J. Chung, Dep. of Agronomy, Agric. College, GyeongSang National Univ., 900 Gaza-Dong, Chinju City, Geyong-Nam, South Korea, 660-701; H.L. Babka, Botany Dep., Iowa State Univ., Ames, IA 50011; G.L. Graef, P.E. Staswick, D.J. Lee, and J.E. Specht, Dep. of Agronomy & Horticulture, Univ. of Nebraska, Lincoln, NE 68583-0915; P.B. Cregan, USDA, ARS, Bldg. 006, Room 100, BARC-West, 10300 Baltimore Ave., Beltsville, MD; R.C. Shoemaker, USDA, ARS, Corn Insect and Crop Genetics Research Unit, Dep. of Agronomy, Iowa State Univ., Ames, IA 50011. Published as Paper no. 13749, Journal Series, Nebraska Agric. Res. Div. Project no. 12-194. Research supported by state and federal funds appropriated to the Agricultural Research Division and University of Nebraska, and by grants received from the United Soybean Board and from Nebraska Soybean Development, Utilization, and Marketing Board. Received 18 June 2002. *Corresponding author (jspecht1@unl.edu).

Published in Crop Sci. 43:1053–1067 (2003).

**Abbreviations:** AFLP(s), amplified fragment length polymorphism(s); BC, backcross; BG, background; bp, base pair; CAPS, cleaved amplified polymorphic sequence; CIM, composite interval mapping; cM, centimorgan; DP, donor parent; LG, linkage group; LOD, base-10 logarithm of the odds ratio (i.e., likelihood of proposed model *versus* null model); LR, linear regression; MG, maturity group; PI, plant introduction; QTL(s), quantitative trait locus (loci); RAPD(s), random amplified polymorphic DNA(s); RFLP(s), restriction fragment length polymorphism(s); RIL(s), recombinant inbred line(s); RP, recurrent parent; SIM, simple interval mapping; SSR(s), simple sequence repeat(s).

correlations between protein and oil are well documented in the soybean genetics and breeding literature (Johnson et al., 1955; Hanson et al., 1961; Hartwig and Hinson, 1972; Shannon et al., 1972; Brim, 1973; Brim and Burton, 1979; Sebern and Lambert 1984; Burton, 1987; Wehrmann et al., 1987; Wilcox and Cavins, 1995; Wilcox, 1998; Cober and Voldeng, 2000).

Seed protein and oil heritabilities are high, especially if the parental differences are extreme (Brim, 1973; Burton, 1987). Selection to enhance either seed constituent is usually successful. Population means and midparent values for protein and oil tend not to be significantly different (Thorne and Fehr, 1970), suggesting an inheritance governed primarily by additive, rather than dominance, genetic effects. Single cross, backcross, and recurrent selection approaches have been used successfully to improve seed protein content (Hartwig and Hinson, 1972; Shannon et al., 1972; Brim and Burton, 1979; Sebern and Lambert, 1984; Wehrmann et al., 1987; Hartwig and Kilen, 1991; Wilcox and Guodong, 1997; Helms and Orf, 1998; Cober and Voldeng, 2000).

In many studies, only a few genes or perhaps even one major gene seemed to govern protein content. Consider, for example, the data of Wilcox and Cavins (1995), who mated the recurrent parent (RP) cultivar Cutler 71 (408 g protein $kg^{-1}$, dry weight basis) to the donor parent (DP) accession Pando (498 g $kg^{-1}$). From that mating, the $F_4$–derived line with the highest seed protein content was selected and backcrossed to the RP, followed by $n = 2$ cycles of again selecting the highest protein BC$n$F$_4$–derived line to backcross to the RP. The $F_4$–derived line selected from the RP × DP, BC1, and BC2 progenies for mating to the RP (i.e., to create the BC1, BC2, and BC3) had near-similar protein contents of 474, 476, and 466 g $kg^{-1}$, respectively. The high protein allele(s) present in the initial (RP × DP) $F_4$–derived line were apparently recaptured after each BC by simply selecting the highest protein BC$n$F$_4$–derived line and using it for the next backcross. The oil content (175 g $kg^{-1}$) of the $F_4$–derived line selected from the RP (204 g $kg^{-1}$) × DP (148 g $kg^{-1}$) mating was also nearly identical to the 174 g $kg^{-1}$ oil content of the final BC3F$_4$ selection. Evidently, low oil allele(s) were cotransmitted with high protein allele(s) between BC cycles by simply advancing the highest protein $F_4$ line to the next backcross. In fact, in each successive backcross progeny, the linear coefficient of regression of protein on oil became more negative and stronger (i.e., a greater $R^2$). Similar results were obtained when Wilcox (1998) subsequently intermated Pando with three cultivars possessing genetic male-sterility, creating a base population for initiating the first of eight cycles of recurrent selection for higher protein. The allele(s) for high protein were mostly fixed by cycle five, such that 66% of the plants in the next cycle produced seed with >479 g $kg^{-1}$ protein, with no significant improvement in later cycles. Again, the inverse relationship between protein and oil steepened and strengthened with each selection cycle. A strongly coordinate high protein *and* low oil response to selection for just higher protein suggests preferential fixation of either (i) high protein allele(s) that were pleiotropic for low oil or (ii) high protein allele(s) linked in repulsion phase with low oil allele(s). If the latter, then such linkage(s) had to be tight enough to preclude the generation of coupling-phased recombinants in populations of the size used in the above studies, otherwise the fixation of any such recombinant would have certainly weakened, if not reversed, the negative correlation between protein and oil.

During the last decade, several research groups, using various population types and different marker-based mapping techniques, have identified many QTLs governing soybean seed protein, oil, and/or yield. Diers et al. (1992) measured the protein and oil content of seed of 60 $F_{2:3}$ lines derived from the mating of a high yielding *G. max* breeding line (A81-356022) with a *G. soja* Siebold & Zucc. accession (PI 468916), then genotyped those 60 lines with 252 marker loci that mapped to 31 linkage groups (LGs). Restriction fragment length polymorphism (RFLP) markers in LG-I (i.e., A144, A407, A688, and K011) and in LG-E (A053, A242, and SAC7) identified a *G. soja* segment in each LG that increased seed protein but decreased seed oil. Sebolt et al. (2000) attempted to confirm these associations by identifying an $F_2$–derived line homozygous for the *G. soja* segments in LG-I and LG-E, and then backcrossing it to the RP (A81-356022). The RFLP marker A144 on LG-I and the classical marker *Pb* on LG-E were used for a marker-assisted introgression of the respective *G. soja* genomic segments into 53 BC3F$_4$–derived lines that were then tested for protein and oil content in multiple environments. The LG-I *G. soja* segment conditioning high protein but low oil was successfully introgressed, but the LG-E segment was not. Lines with the introgressed LG-I segment exhibited earlier maturity, taller plants, smaller seed, and lower yields. A BC3F$_4$–derived line homozygous for the LG-I *G. soja* segment was then mated to the cultivars Parker, Kenwood, and C1914, the latter a high protein breeding line of Pando origin. Segregation of the LG-I *G. soja* segment was readily detectable in the Parker and Kenwood $F_2$ populations (i.e., 10 g more protein but 4.5 g less oil per kg of seed), but not in the C1914 $F_2$ populations. That result led Sebolt et al. (2000) to suggest that the LG-I high protein allele of the *G. soja* accession (PI 468916) was probably allelic with the high protein allele of C1914 (and by inference, Pando).

The LG-I protein (or oil) QTL has since been detected by others. Brummer et al. (1997) examined eight different populations of $F_2$–derived lines and identified associations of soybean protein and/or oil content with RFLPs of various linkage groups. One of the strongest associations was with RFLPs A407 and A144 in a population derived from the mating of a breeding line (M82-806) with a high protein line (HHP; 25% *G. soja* by pedigree). Mansur et al. (1993), using an $F_5$ population from a 'Minsoy' × 'Noir 1' mating, reported a protein QTL associated with RFLP L048. No association of any LG-I marker with protein was detected in a recombinant inbred line (RIL) population of the same mating (Mansur et al., 1996), although Specht et al. (2001) did report an oil QTL between LG-I SSRs Satt127 and

Satt239. Csanádi et al. (2001) detected a LG-I QTL for oil (and for protein in some tests) in a mating of the early-maturing cultivars Proto (462 g $kg^{-1}$ protein) and Ma Belle (390 g $kg^{-1}$ protein). Lee et al. (1996), Orf et al. (1999), and Qiu et al. (1999) did not detect a LG-I QTL for protein or oil, but protein variation in their populations was small. Statistical parameters for the protein and oil QTLs detected by each of the foregoing researchers were recently summarized (see Appendix Table I of Olsen, 2001). Map positions of the reported protein and oil QTLs are also available on-line (SoyBase, 2002).

We report here the results of a QTL analysis of seed protein, oil, and yield in a population of 76 $F_5$–derived RILs from a high protein *G. max* germplasm accession PI 437088A mated to a high yielding *G. max* cultivar Asgrow A3733. Surprisingly, the RILs segregated for the same LG-I protein QTL that has been detected in other *G. soja* and *G. max* germplasm. Because this LG-I QTL seems to be more ubiquitous in high protein germplasm than originally surmised, our objective in this research was to delineate more precisely the map position of this LG-I QTL by using more markers than had been used in prior studies. Our second objective was to obtain more precise estimates of the direction and magnitude of the parental allele effects on RIL seed protein, oil, and yield, and to interpret these allelic effects with respect to the observed genotypic-level correlations among those three traits.

## MATERIALS AND METHODS

### Parental Germplasm and Population Development

The *G. max* accession PI 437088A was collected in a "far eastern" region of the former USSR by the N.I. Vavilov Institute of Plant Industry of the Russian Federation and donated in 1979 to the USDA Soybean Germplasm Office (NGRP, 2001). PI 437088A is a low yielding maturity group I accession whose seed is high in protein but low in oil (>480 g $kg^{-1}$ and <160 g $kg^{-1}$ seed, respectively, dry weight basis). Asgrow A3733 is a high-yielding MG III proprietary cultivar with a seed protein and oil content similar to that of most high-yielding cultivars grown in the north central USA (420 g $kg^{-1}$ and 205 g $kg^{-1}$).

PI 437088A was mated as a male to Asgrow A3733 in the summer of 1992. Five $F_1$ plants were grown in the greenhouse during the winter of 1992-93. The $F_2$ generation was grown in the summer of 1993. The $F_3$ and $F_4$ generations were grown at the USDA Tropical Agriculture Research Station at Isabela, Puerto Rico, during the November to February and February to May seasons of 1993-1994, and advanced from the $F_2$ to $F_5$ by single-seed descent. The $F_5$ generation was grown at Lincoln, NE, during the summer of 1994. About 100 random $F_5$–derived $F_6$ (i.e., $F_{5:6}$) progenies were grown at Lincoln, NE, in 1995 for a seed increase, but only 76 generated sufficient seed for inclusion in the subsequent field trials.

### Quantitative Trait Measurement and Analysis

Performance trials of the 76 RILs were conducted in the $F_{5:7}$ (1996) and $F_{5:8}$ (1997) generations at the Agricultural Research and Development Center near Mead, NE. The soil at the test site is a Sharpsburg silty clay loam (fine, smectitic, mesic Typic Argiudoll). Each year, the experimental design was a randomized complete block with two replicates. The treatment design was a split-plot. The main plots were six seasonal water amounts cumulatively generated by weekly sprinkler irrigation that replaced 0, 20, 40, 60, 80, 100% of the evapotranspiratory water loss (adjusted for rainfall) since the prior irrigation, a procedure described by Specht et al. (1986, 2001). The 80 subplots consisted of 76 RIL entries plus two entries of each of the two parents. Each subplot was comprised of two 3.05-m plant rows spaced 0.76 m apart. The field trials were planted on 30 May 1996 and 13 May 1997 at a seeding rate of 370 000 viable seed $ha^{-1}$ and a sowing depth of 3 cm.

The two-row subplots were harvested with a self-propelled plot combine to estimate subplot seed yield (kg $ha^{-1}$). Two 100-seed samples taken from each subplot were used to estimate 100-seed weight (g). Seed yield and 100-seed weight data were adjusted to the standard 130 g $kg^{-1}$ seed moisture content. A 75-g seed sample randomly drawn from each subplot was assayed with the standard near-infrared transmittance technique to quantify seed protein and oil content (g $kg^{-1}$) on a zero percentage moisture basis (Panford, 1987). Maturity (d from planting), plant height (cm from ground to stem tip), and plant lodging (visually scored on a scale of 1 = erect to 5 = prostrate) data were collected before harvest, but only on the 0 and 100% irrigated subplots.

Trait data were subjected to an analysis of variance and covariance by means of the PROC GLM module (with a MANOVA statement) of version 8 of PC SAS for Windows (SAS, 1999). Years, water levels, genotypes, and replicates were treated as random effects. Heritability was estimated on an entry mean basis. The associated confidence intervals were computed as described by Knapp et al. (1985). Genotypic correlations between traits were estimated as described by Mode and Robinson (1959).

### Leaf Collection and DNA Isolation Procedures

In early July of 1996, young leaf tissue was collected from 25+ plants of each RIL (and each parental) plot of the first replicate, transported to the laboratory on dry ice, frozen, lyophilized, and then ground to a fine powder before being stored at −20°C. DNA was extracted by means of a protocol modified from that described by Saghai-Maroof et al. (1984). After extraction of DNA and its resuspension in TE, the DNA samples were diluted to 20 ng $\mu L^{-1}$ for marker analysis.

### Molecular Marker Types and Assay Protocols

The initial genomic characterization of the parents and RILs was performed by means of RAPD markers and a polymerase chain reaction (PCR) protocol (Williams et al., 1990). Each PCR reaction was performed in 10 μL containing 50 m*M* Tris (pH 8.5), 2 m*M* $MgCl_2$, 20 m*M* KCl, 0.5 mg $mL^{-1}$ bovine serum albumin (BSA), 2.5% (w/v) Ficoll 400, 0.02% (w/v) xylene cyanol, 4 ng $\mu L^{-1}$ template DNA, 0.4 μ*M* primer, 100 μ*M* of each dATP, dCTP, dGTP, and dTTP, and 0.6 unit Taq polymerase. To prevent evaporation during PCR, the reaction mixture was covered with a mineral oil drop. PCR was conducted in a thermocycler (Model PT-100, MJ Research, Watertown, MA) programmed to this three-step cycling protocol: (i) two cycles in which the reaction mix tubes were heated to 91.0°C for 75 s to allow DNA denaturation, cooled to 42.0°C for 22 s to allow primers to bind, then reheated to 72.0°C for 70 s to allow DNA synthesis by the polymerase; (ii) 38 cycles of 16 s at 91.0°C, 2 s at 42.0°C, and 70 s at 72°C; (iii) one final cycle in which the extension step lasted 4 min at 72.0°C, before the reaction mixes were cooled to 4°C. The amplification products were loaded on 1.2% (w/v) agarose gels (Amresco, 3:1

agarose, Solon, OH) for electrophoretic separation at 74 V for 4 h. Ethidium bromide staining of the separated amplicons allowed their visualization under UV light. All scored amplicons were well within the 2176- to 154-base pair (bp) sizes of largest and smallest markers of the molecular weight standard VI (Boehringer Mannheim, Indianapolis, IN).

About 1000 10-bp primers differing in sequence were obtained from Operon Technologies (Alameda, CA) to screen the parents for amplicon presence (+)/absence (−), but only 370 of the primers produced an amplicon bimorphism. These primers were used to assay 20 random RILs to determine if the parental +/− amplicon difference actually segregated. About 340 amplicons met this test and were then examined in all 76 RILs. Eleven amplicons whose +/− segregation in the initial screen was not reproduced with complete fidelity in the final screen were dropped. The remaining 329 amplicons were then assigned a marker name consisting of the Operon product code for the primer (e.g., OP_A01), but appended with a small-case letter suffix if more than one +/− amplicon was produced by the primer (e.g., OP_A01a, OP_A01b, etc.).

Codominant SSR markers were not publicly available in sufficient numbers until late 1998 (Cregan et al., 1999). Thereafter, the two parents were screened with about 250 SSR primers. Only 103 of these were parentally polymorphic, but were distributed over the 20 linkage groups. The SSR-PCR amplification protocols were similar to those used by Akkaya et al. (1992) and Rongwen et al. (1995). The PCR reaction mix contained 50 ng of genomic DNA (parent or RIL), 0.1 $\mu M$ of each of the paired primers, 5× reaction buffer (250 m$M$ Tris), 0.6 unit DNA polymerase, and 2.5 $\mu M$ of each of the four dNTPs. Each sample was subjected to 31 cycles of denaturing (25 s at 94°C); annealing (25 s at 47°C); and extension (25 s at 68°C) in a thermocycler, followed by a final extension step (3 min at 72°C) and incubation at 4°C. For 74 of the 103 SSRs, electrophoretic separation of their $^{32}$P-labeled amplicons on polyacrylamide gels was contracted to Biogenetics Services Inc. (Brookings, SD). Parental amplicons of the other 29 SSRs were sufficiently different in base pair size to permit their electrophoretic separation on high-resolution 5% (w/v) agarose gels (SFR Agarose, Amaresco, Solon, OH) at 74 V for 4 h, followed by visualization via ethidium bromide staining. The SSR amplicons fell within the 2642- to 50-bp size range of the markers in molecular weight standard XIII (Boehringer Mannheim). The 20- to 24-bp sequences of the 103 SSR primer pairs are available on-line (SoyBase, 2002).

## Classical Markers

The 76 RILs segregated for the black/brown seed hilum color phenotypes (on a yellow seed coat) governed by the *R/r* locus on linkage group K (LG-K), and for the brown/tan pod color phenotypes governed by the *L2/l2* locus on LG-N (Shoemaker and Specht, 1995). The genotypes of the PI 437088A and Asgrow A3733 parents were thus *rrl2l2* and *RRL2L2*, respectively.

## Cysteine Protease CAPS Marker

Endoproteolytic cleavage has been associated with the accumulation of soybean storage proteins (Herman and Larkins, 1999). However, the two parents were not polymorphic for the endopeptidase isozyme locus (*Enp*) located on LG-I. A search of GenBank (NCBI, 2001) indicated that only one soybean endopeptidase had been cloned to date—a cysteine protease gene (GenBank Acc. No. D28876). To map this gene, a cleaved amplified polymorphic sequence (CAPS) marker was constructed. An oligonucleotide primer pair was synthesized to amplify a unique portion of the published cDNA sequence. When parental PCR products were tested with a battery of restriction enzymes, a *Hin*fI restriction site was detected in the Asgrow A3733 amplicon but not in the PI 437088A amplicon. The 888-bp uncleaved *Hin*fI fragment from PI 437088A was subcloned and sequenced to construct the following forward and reverse primers:

PIHNF1: CAACTTACAAGATGCTCCG
PIHNF2: ATCATCAACCACCCTGAG

This primer pair amplified an 844-bp fragment in each parent. On treatment with *Hin*fI, the Asgrow 3733 amplicon was cleaved into two fragments (580 and 264 bp), whereas the PI 437088A amplicon remained intact. The PIHNF- generated amplicons from each of the 76 RILs, after *Hin*fI treatment, were loaded on 1.2% agarose gels (Amresco, 3:1 agarose, Solon, OH), electrophoresed at 74 V for 4 h, then stained with ethidium bromide for visualization under UV light. This codominant CAPS marker was assigned the marker name CystProt.

## Mature Seed Protein Marker

Because the two parents differed substantially in seed protein content, a gel-based search for qualitative protein differences in mature parental seed was conducted. A 50-seed sample, taken from pure RIL and parent seed stock produced in a common environment, was ground into a fine powder that was then stored at −20°C. A sample of soybean flour (i.e., a microcentrifuge tube filled with flour to the 0.5-μL line) was placed into a mortar (on ice) to which 1 mL of cold extraction buffer was added. The extraction buffer consisted of 50 m$M$ Tris (pH 7.5), 10 m$M$ EDTA (pH 8.0), 5 m$M$ dithiothreitol (DTT), 1% (w/v) insoluble polyvinylpolypyrrolidone (PVPP), and 0.5 m$M$ phenylmethylsulfonyl fluoride (PMSF). The flour and buffer were mixed in the mortar with the pestle until a thick slurry formed, which was then poured into a 1.5-mL microcentrifuge tube and spun in a centrifuge at 16 000 *g* for 10 min at 4°C. The supernatant was removed and its protein concentration was determined (Bio-Rad DC Protein Assay, Hercules, CA). The protein samples were then diluted to 3 μg $\mu L^{-1}$, mixed with the SDS-PAGE loading buffer, denatured by boiling for 5 min, then cooled. Sample aliquots of 30 μL were loaded into individual wells of a 12% polyacrylamide resolving gel (Bio-Rad, Hercules, CA), with 1× TG buffer in the upper chamber and 0.5× TG buffer in the bottom chamber. Electrophoresis was conducted at 25 mA for about 4 h, when the dye front moved off the gel edge. Each parent exhibited a unique protein band the other lacked, but the two bands segregated in the RILs as codominant alleles at a single locus that was assigned the marker name MatSdPro.

## Linkage Map Construction

The observed allelic segregation ratios at each of the 436 loci—329 RAPDs, 103 SSRs, two classical markers, one CAPS marker, and one seed protein marker—were subjected to Chi-square tests for goodness-of-fit to expected ratios. For codominant SSR markers, a 15A:2H:15B segregation ratio was expected in this $F_5$–derived set of RILs, where the letters code for RILs homozygous for the A allele of Asgrow A3733 or the B allele of PI 437088A, or for heterozygous AB RILs. For dominant RAPD markers, the expected segregation ratio was either 17D:15B, with D coding for the indistinguishable AA and AB RILs when the B allele produced no amplicon, or 17C:15A, with C coding for the indistinguishable BB and AB RILs when the A allele produced no amplicon. To obtain

a genomewise significance criterion of $\alpha' = 0.05$ for this large number (436) of 3- or 2-class Chi-square tests, an approximate Bonferroni adjustment was applied to ensure an appropriate testwise significance criterion (i.e., $\alpha = \alpha'/436 = 0.0001$).

Mapmaker/exp 3.0 software (Lander and Botstein, 1989; Lincoln and Lander, 1993) was used to create a genetic map of the *ri self* type, which required replacing the few SSR marker H (heterozygote) codes in the raw data file with dashes (e.g., missing data). Mapmaker's Group command, used with a stringent grouping LOD of 5.5 to preclude spurious grouping results, led to 385 of the 436 markers coalescing into 50 LGs. Each of those 50 LGs was then tested one at a time for linkage to the other 49 other LGs by means of a relaxed grouping LOD of 3.0. This resulted in an unambiguous fusion of 15 LGs with other LGs, reducing to 35 the total number of LGs. Those 35 LGs were then tested one at a time for linkage with each of the 51 unlinked markers, also by means of a relaxed grouping LOD of 3.0. This resulted in the unambiguous linkage of 31 of the 51 markers with the 35 LGs, leaving 20 markers (of the 436) not linking to any of the 35 LGs. Marker ordering within LGs of eight or fewer markers was accomplished directly by use of Mapmaker's Compare command. Marker ordering within the larger LGs was accomplished by Mapmaker's Order command. A "seed order" was established by means of a minimum grouping LOD of 3.0, a minimum seed order start size of five markers, and a minimum distance of 2 cM between those markers. Each remaining marker was then placed in that order by LOD placement thresholds of 4.0 (strict) and 3.0 (relaxed), and a minimum window size of 3 (i.e., at least one marker on each side of a placed marker). After all markers were placed in the LG, the Ripple command was then used with a LOD threshold of 3.0 and a (marker) window size of 6 to identify/correct any minor marker placement errors.

### QTL Analyses

QTL analysis was performed by the linear regression (LR), simple interval mapping (SIM), composite interval mapping (CIM) modules of QTL Cartographer (Zeng, 1994; Basten et al., 2001). The LOD score criteria for evaluating the statistical significance of QTL effects in each module were estimated by permutation (Churchill and Doerge, 1994). A minimum of 1000 permutations was generated in each module (i.e., LR, SIM, and CIM) for each trait–year data set. Permutation output differences among the trait-year data sets were small, so they were averaged to obtain final LOD score criteria (equivalent to $\alpha' = 0.05$ genomewise error rates) of 3.4 for LR, 3.2 for SIM, and 3.8 for CIM analyses. A limited number of background (BG) markers for the CIM analysis were identified via the forward/backward stepwise regression option of QTL Cartographer using conservative probability thresholds (*Pin* = 0.01; *Pout* = 0.01). A CIM window parameter of 1 cM was chosen to exclude, from the BG marker group, any marker located within 1 cM of the two markers flanking an interval being tested for a putative QTL peak.

The joint map analysis module of QTL Cartographer was applied to the 1996 and 1997 data of each trait to evaluate the significance of G × E interaction, which in the present case is actually the QTL × Y interaction for each trait. Jiang and Zeng (1995) noted the interval maximum for the joint mapping statistic is roughly approximated by a Chi-square distribution with $2m+1$ degrees of freedom, where $m$ = the number of jointly mapped traits (i.e., $m = 2$ for the 1996 and 1997 values of the trait). To derive a genomewise error rate of $\alpha' = 0.05$ for the joint trait mapping statistic, the Bonferroni correction was applied: $\alpha = 1 - (1 - \alpha')^{1/M}$, where $M$ is the total number of marker intervals in the genome. In the present case, $M = 381$ (i.e., 416 loci in 35 LGs, see Fig. 1). Thus, $\alpha = 0.000134619$ which, with $2\times2+1 = 5$ degrees of freedom, translated into a Chi-square value of 25.08, which was equivalent to a LOD score of 5.4 ($25.08 \times 0.217$). The G × E statistic itself was computed *only* for those few marker intervals whose joint mapping LOD statistic exceeded the 5.4 criterion, and it is asymptotically distributed as a Chi-square with two degrees of freedom (Jiang and Zeng, 1995). Thus, at $\alpha = 0.05$, the G × E statistic has a Chi-square significance criterion of 5.99, which translates into a LOD score of 1.30 ($5.99 \times 0.217$).

Jiang and Zeng (1995) have described a joint mapping method for statistically evaluating pleiotropy vis-à-vis tight linkage, but this method has not yet been implemented in the recent version of QTL Cartographer. In the absence of a priori evidential disproof, pleiotropism is usually treated as the null hypothesis since it is disprovable on the occurrence of a confirmed recombinant (Hanson, 1959). In the present study, pleiotropism was assumed if (i) a QTL for one trait and a QTL for another trait had a coincident map position, or (ii) if the confidence interval for the QTL of one trait encompassed the mean map position of the QTL of another trait.

## RESULTS

### Genotypic Data and Linkage Map

Segregation data were obtained for 436 genomic loci in this population of 76 RILs. These loci included 329 RAPDs, 103 SSRs, one CAPS marker, one seed protein mobility variant, and two genes for pod and seed pigmentation. More than 95% of these loci (i.e., 416) coalesced into 35 LGs (Fig. 1), which, on the basis of SSR marker homology, corresponded in whole or in part with the known and named 20 soybean LGs (Cregan et al., 1999). The reductional alignment of 35 LGs to 20 LGs resulted in 15 linkage gaps. A pair of repulsion-phased dominant (RAPD) markers flanked nine of those gaps, with a dominant (RAPD)–codominant (SSR) marker pair flanking five other gaps. Linkage is difficult to estimate in the case of loosely linked, repulsion-phased dominant markers because of minimal information content and low statistical power (Allard, 1956; Skroch and Neinhuis, 1995; Liu, 1998). Linked pairs of dominant and codominant markers also have substantially less information content than linked pairs of two codominant markers. Moreover, heterozygotes are retained in an $F_5$–derived RIL population at a frequency of 1/16 at each marker locus. While heterozygotes at codominant SSR marker loci were readily identified, RAPD marker heterozygotes were not distinguishable from, and were thus unavoidably misclassified with, their respective dominant homozygotes (i.e., 15A+2H: 15B or 15A: 2H+15B). Thus, map distances that include RAPD markers may be somewhat inflated.

Twenty of the 436 markers failed to group into LGs. Fourteen (13 RAPDs and one SSR) had significantly distorted segregation ratios. The six others (all SSRs) occupy terminal map positions on LGs (Cregan et al., 1999) and did not link because the linkage distance to the next inward marker was >45 cM. These terminal SSRs included Satt371 at the bottom of LG-C2, Satt184 at the top of LG-D1a, Sctt008 at the top of LG-D2,

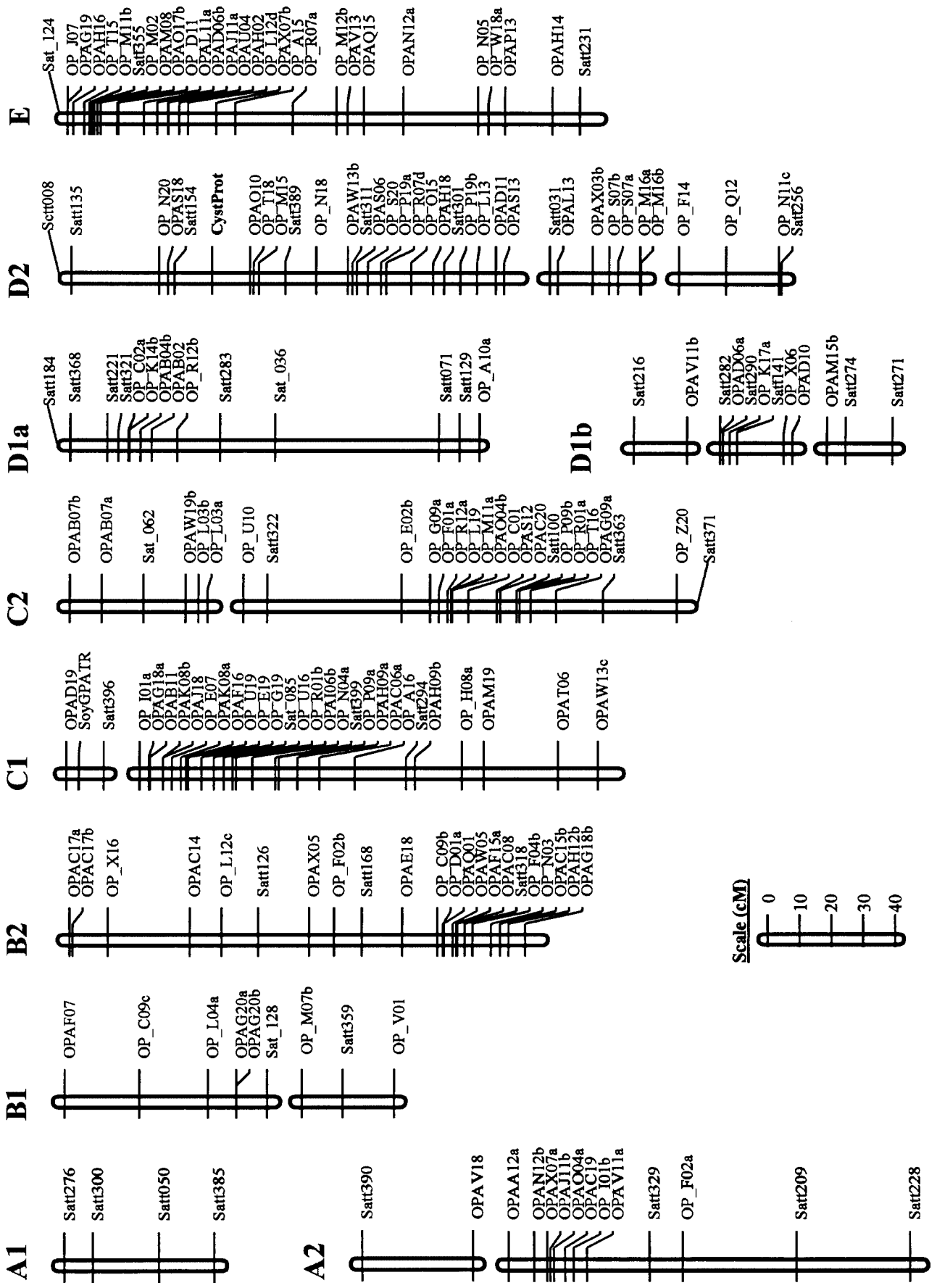
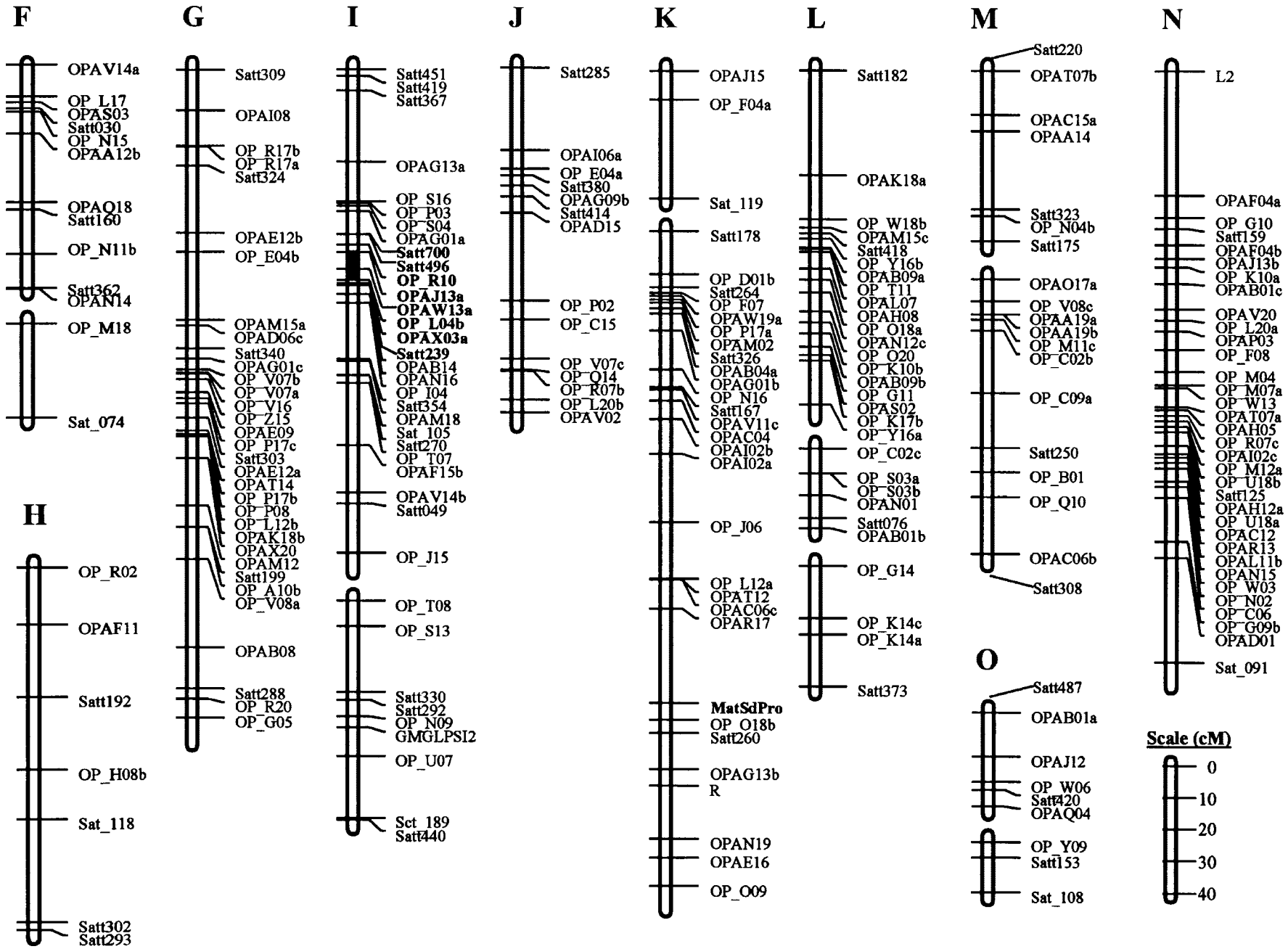



**Fig. 1. A genetic linkage map derived from a marker analysis of 76 $F_{5:6}$ recombinant inbred lines of the Asgrow A3733 × PI 437088A mating. The 417 markers of this map coalesced into 35 linkage groups which were matched, on the basis of SSR marker homology, to the 20 known and named A through O soybean linkage groups (Cregan et al., 1999). Numerical marker distances were deleted to minimize figure size, but**




all linkage groups are shown on the same proportionate scale. The Haldane mapping function (centimorgan units) was used to create the map. The small solid rectangle in linkage group (LG) I denotes the map position of the protein, oil, and yield QTL(s).

**Table 1. Parental means and population statistics for seed protein, oil, and yield. The $F_5$–derived population of 76 recombinant inbred lines (RILs) originated from the mating of Asgrow A3733 × PI 437088A. In 1996 and 1997, the parents and 76 RILs were evaluated in two replicates of six irrigation treatments (i.e., 12 observations per RIL mean in each year).**

| | 1996 Seed | | | 1997 Seed | | |
|---|---|---|---|---|---|---|
| **Parameter** | **Protein** | **Oil** | **Yield** | **Protein** | **Oil** | **Yield** |
| | **10 g kg$^{-1}$** | | **kg ha$^{-1}$** | **10 g kg$^{-1}$** | | **kg ha$^{-1}$** |
| **Parental means** | | | | | | |
| **Asgrow A3733** | **41.8** | **20.8** | **3834** | **42.5** | **20.4** | **3441** |
| **PI 437.088A** | **48.7** | **15.9** | **1337** | **47.9** | **16.2** | **2111** |
| **Population statistics** | | | | | | |
| **Mean** | **44.8** | **18.6** | **2998** | **45.4** | **18.0** | **2641** |
| **Minimum** | **41.3** | **16.1** | **2429** | **41.8** | **15.2** | **1938** |
| **Maximum** | **48.2** | **20.4** | **3882** | **49.4** | **20.7** | **3513** |
| **Standard deviation** | **1.4** | **1.0** | **286** | **1.6** | **1.1** | **389** |
| | | | **No units** | | | |
| **Population distribution** | | | | | | |
| **Kurtosis** | **−0.39** | **−0.34** | **+0.19** | **−0.50** | **−0.23** | **−0.75** |
| **Skewness** | **+0.11** | **−0.50** | **+0.22** | **+0.14** | **−0.17** | **+0.29** |
| **Normality test statistic †** | **0.36** | **3.31** | **0.92** | **0.70** | **0.38** | **2.34** |

**† A test statistic of less than 5.99 ($\alpha = 0.05$) indicated the trait data were normally distributed.**

Sat_124 at the top of LG-E, Satt308 at the bottom of LG-M, and Satt487 at the top of LG-O (Fig. 1).

The summed genomic distance of 2943 cM for the RIL map in Fig. 1 compared favorably with the 3003-cM map of RFLP–SSR markers in the $F_7$–derived RIL population documented in Cregan et al. (1999). Yamanaka et al. (2001) recently reported a 2909-cM map based on RFLP–SSRP markers in a $F_2$ population. Ferreira et al. (2000) reported a 3275-cM map of 106 RAPD and 250 RFLP markers in an $F_6$–derived RIL population. Keim et al. (1997) reported a summed distance of 3441 cM for a map of 840 mostly dominant AFLP markers in the $F_6$–derived RIL population. In the latter two reports, marker clustering was evident in the maps. The authors speculated that this clustering was likely due to markers occupying positions in heterochromatic (possibly centromeric) regions of the chromosomes in which recombination was greatly reduced. Some clustering of RAPD markers was also evident in our map (Fig. 1).

The observed segregation of the codominant CAPS marker CystPro, and that of the codominant protein mobility marker MatSdPro, did not differ significantly from the expected 15A:2H:15B ratio. The CystProt marker mapped to LG-D2, about 12.1 cM from Satt154, whereas the MatSdPro marker mapped to LG K, about 10 cM from Satt260 (Fig. 1).

## Phenotypic Data

The analyses of variance revealed that the 0 to 100% irrigation treatments had a modest, though statistically significant, linear impact on all six traits. Seed yield, seed protein, 100-seed weight, maturity, and plant height were enhanced, whereas seed oil was depressed by incrementally greater amounts of seasonal water. Such responses have been observed in prior studies (Specht et al., 1986, 2001). The irrigation × RIL and the year × irrigation × RIL interactions were not significant for any trait. Although the year × irrigation interaction was significant for seed protein and oil, it was simply due to a larger linear effect of irrigation in 1996 than in 1997. These findings indicated that RIL means could be averaged over irrigation treatments.

Significant differences among the RILs were detected for all traits. The RIL × year interaction was significant for yield, protein, oil, and 100-seed weight, indicating that RIL means could not be pooled over years for subsequent QTL analyses. This interaction was not unexpected, particularly for yield, and might have been attributable to a *Bean pod mottle virus* epidemic that occurred in 1997. One symptom of viral infection was the failure of plant stems to senesce at physiological maturity, resulting in what is commonly termed "green stems" (Schwenk and Nickell, 1980). Pods on green stems eventually matured, but the green stems, when crushed by the combine cylinder, exuded moisture onto seeds and pods, making difficult a clean separation of seed from haulm (i.e., pod walls, stems and branches, plus any nonabscised petioles and leaves). This led to imprecision in the 1997 yield measurement. Variable delays in pod maturity also rendered unreliable the estimates of 1997 RIL maturity.

In both years, the seed protein content of the PI 437088A parent substantially exceeded that of the Asgrow A3733 parent, whereas the inverse was true for seed oil and yield (Table 1). The population means were near the mid-parent values. The trait data were normally distributed.

Heritabilities on an entry mean basis were computed for seed protein, oil, and yield for single-year and 2-yr data (Table 2). The 2-yr protein and oil heritabilities computed for this population of 76 RILs were near the upper limit of those cited in the Brim (1973) and Burton (1987) reviews of soybean quantitative genetic parameters. This was probably because the error variances for the seed constituent traits in the present study were quite small. In fact, the 1996 and 1997 coefficients of variation were 1 and 2% for seed protein and 2 and 4% for seed oil, compared with 8 and 12% for seed yield, respectively. Moreover, all interaction variances were small relative to the genetic variances, including the statistically significant RIL × year interaction variances.

The genotypic correlations among yield, protein, and oil were quite similar between years (Table 3). The exceptionally negative genotypic correlation of protein

**Table 2. Heritabilities on an entry mean basis and their associated confidence intervals (CI) for soybean seed protein, oil, and yield measured in the $F_5$–derived population of 76 recombinant inbred lines (RILs) from the mating of Asgrow A3733 × PI 437088A. In 1996 and 1997, the 76 RILs were evaluated in two replicates of six irrigation treatments (i.e., 12 observations per RIL mean in each year). For all three traits, the only significant RIL interaction was that of RIL × year.**

| | 1996 Seed | | | 1997 Seed | | | 1996–97 Seed | | |
|---|---|---|---|---|---|---|---|---|---|
| Parameter | Protein | Oil | Yield | Protein | Oil | Yield | Protein | Oil | Yield |
| Upper 95% CI | 0.992 | 0.989 | 0.972 | 0.992 | 0.990 | 0.983 | 0.931 | 0.895 | 0.769 |
| Heritability ($h^2$) | 0.988 | 0.985 | 0.961 | 0.988 | 0.986 | 0.977 | 0.889 | 0.844 | 0.653 |
| Lower 95% CI | 0.983 | 0.978 | 0.943 | 0.983 | 0.979 | 0.967 | 0.833 | 0.745 | 0.444 |

and oil ($r > -0.8$) was not surprising, given the well-known difficulty of achieving tandem genetic improvement in *both* protein and oil content at the expense of carbohydrate content (Hanson et al., 1961; Burton, 1987). The corresponding environmental correlation ($r > -0.8$) was just as strong, indicating that alteration of protein content by a nongenetic factor coordinately induced a reciprocal alteration in oil content. The genotypic correlations of yield with protein, and yield with oil, were nearly equivalent in magnitude, but opposite in sign. The same was true of the corresponding environmental correlations, though these were much smaller in magnitude. Reports of positive genotypic correlations for yield and oil, but negative correlations for yield and protein, are commonplace in the literature (Johnson et al., 1955; Hanson et al., 1961; Thorne and Fehr, 1970; Hartwig and Hinson, 1972; Shannon et al., 1972; Brim, 1973; Brim and Burton, 1979; Sebern and Lambert, 1984; Burton, 1987; Wehrmann et al., 1987; Hartwig and Kilen, 1991; Wilcox and Cavins, 1995; Wilcox and Guodong, 1997; Helms and Orf, 1998; Wilcox, 1998; Cober and Voldeng, 2000).

RIL seed protein and oil values were plotted against RIL seed yield values in the 1996 and 1997 field trials (Fig. 2). The phenotypic linear regression coefficients were negative when protein was regressed on yield (i.e., $b = -2.86\%$ per Mg ha$^{-1}$ in 1996; $b = -2.34\%$ per Mg ha$^{-1}$ in 1997), but were positive when oil was regressed on yield (i.e., $b = +1.62\%$ per Mg ha$^{-1}$ in 1996; $b = +1.55\%$ per Mg ha$^{-1}$ in 1997). Note that these coefficients reflected the change in seed protein or oil that occurred per unit of change in yield. Thus, a yield increase of 1.0 Mg ha$^{-1}$ was coordinately associated with a 2.34 to 2.86 percentage point depression of seed protein, and a 1.55 to 1.62 percentage point enhancement of seed oil. Note also that the protein/oil exchange ratio per unit of yield change was −1.77 (i.e., −2.86/+1.62) in 1996 and −1.51 (i.e., −234/+155) in 1997. Thus, for each unit of yield increase, 1.51 to 1.77 units of seed protein were exchanged for 1.0 units of oil, with the inverse occurring for each unit of yield decrease. In a study conducted more than 40 yr ago, Hanson et al. (1961) reported protein/oil exchange ratios of −1.6 to −1.7.

In the present study, the 1996 and 1997 coefficients derived from the regression of protein on oil were $b = -1.16$ and −1.26, while those derived from oil regressed on protein were $b = -0.58$ and-0.64 (Fig. 2). These values translate into a protein/oil exchange ratio without reference to yield of 2.00 ( −1.16/+0.58) and 1.97 (−1.26/+0.64). Hanson et al. (1961) reported a protein/oil exchange ratio of 1.92 at constant yield, and noted that this value was quite close to the expected 2.0 ratio of the (average) calorific values reported for soybean oil versus protein.

## QTL Analysis

Statistically significant QTLs for soybean seed protein and oil in both years, and for yield and maturity in 1996, were detected in a genomic region located in the upper half of LG-I (Fig. 1), but were not detected elsewhere in the soybean genome (except for 1996 yield and maturity). No significant QTL for 1997 yield was detected anywhere in the genome, probably because of yield variability associated with the difficulty of completely threshing the green-stemmed plants that year. Statistically significant 1996 and 1997 QTLs were detected for plant height, lodging, and 100-seed weight, but none of these QTLs mapped to LG-I (data not shown), so these will not be discussed in this paper. The two parentally allelic seed protein markers, CystPro on LG-D2 and MatSdPro on LG-K, were not associated with the parental difference in seed protein content, since no statistically significant QTLs for seed protein content were detected near those markers.

The LG-I QTL scans generated by the LR, SIM, and CIM modules of QTL Cartographer are presented in Fig. 3. The LOD peaks and valleys detected in the LR and SIM scans suggested that LG-I possessed one major QTL and perhaps two or three (nearby) minor QTLs for protein, oil, and 1996 yield. Other researchers using

**Table 3. Genotypic correlations (upper right diagonal) and partial correlations derived from the error covariance matrix (lower left diagonal) between soybean seed protein, seed oil, and seed yield in an $F_5$–derived population of 76 recombinant inbred lines (RILs) from the mating of Asgrow A3733 × PI 437088A. In 1996 and 1997, the 76 RILs were evaluated in two replicates of six irrigation treatments (i.e., 12 observations per RIL mean in each year). For all traits, the only significant RIL interaction was that of RIL × year.**

| | 1996 Seed | | | 1997 Seed | | | 1996–1997 Seed | | |
|---|---|---|---|---|---|---|---|---|---|
| | Protein | Oil | Yield | Protein | Oil | Yield | Protein | Oil | Yield |
| Protein | – | −0.824 | −0.582 | – | −0.899 | −0.572 | – | −0.902 | −0.695 |
| Oil | −0.844 | – | +0.465 | −0.808 | – | +0.531 | −0.813 | – | +0.459 |
| Yield | −0.029 | +0.094 | – | −0.230 | +0.360 | – | −0.171 | +0.278 | – |

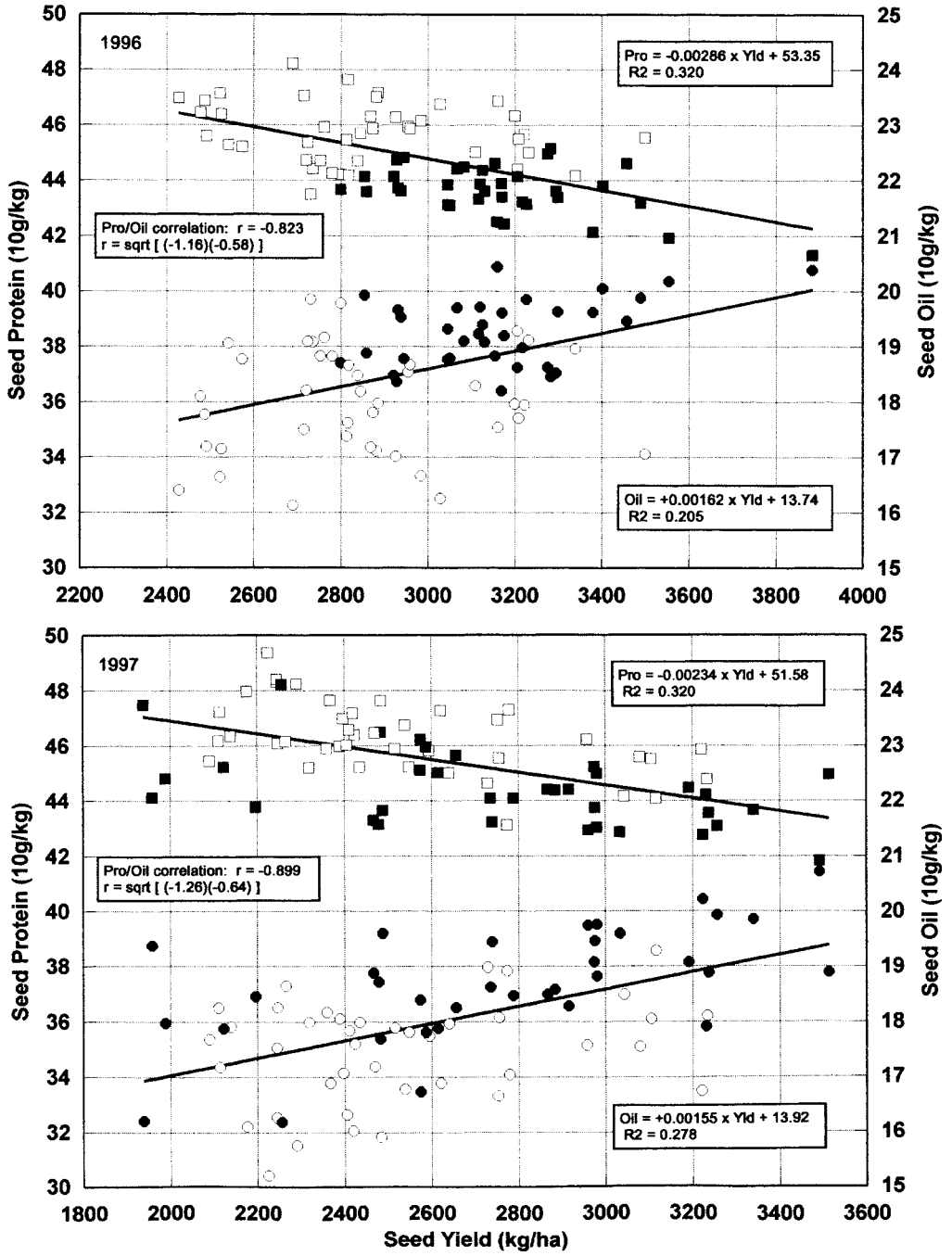



**Fig. 2. A graph of seed protein (left axis- square symbols) and seed oil (right axis-circle symbols) versus seed yield (bottom axis) of each of the 76 $F_5$–derived recombinant inbred lines (RILs) in 1996 and 1997. The solid and open symbols represent RILs with a respective AA (Asgrow A3733) or BB (PI 437.088A) parental genotype at the RAPD marker locus OPAW13a. Trend lines and *b* coefficients in the right text boxes were generated by linear regression of RIL protein or oil on the yield of the same RIL. The *b* coefficients of protein regressed on oil, and oil regressed on protein, are shown in the left text box, and the square root of product of those *b* coefficients is the correlation of protein with oil.**

SIM have reported multiple QTLs in LG-I, and in fact SOYBASE (2002) currently lists 11 protein QTLs (numbered 1-1 to 1-8, 3-12, 10-1, and 11-1) in this small region of LG-I. However, CIM evaluates marker-flanked genomic intervals for evidence of a QTL with greater precision than SIM, by using background (BG) markers to control trait variation arising from the BG genomic segments residing outside of each (window-sized) marker-flanked segment being tested for a QTL (Zeng, 1994). To avoid the well-known problem of identifying too many BG markers for CIM, conservative threshold parameters (i.e., *Pin* = *Pout* = 0.01) were used in this study for the forward (in), backward (out), stepwise regression module of QTL Cartographer. The window size parameter was also narrowed to 1 cM (from its default of 10 cM) to enable a more precise positioning of the protein, oil, and yield QTLs on LG-I, and to ensure a more definitive estimate of each allelic effect.

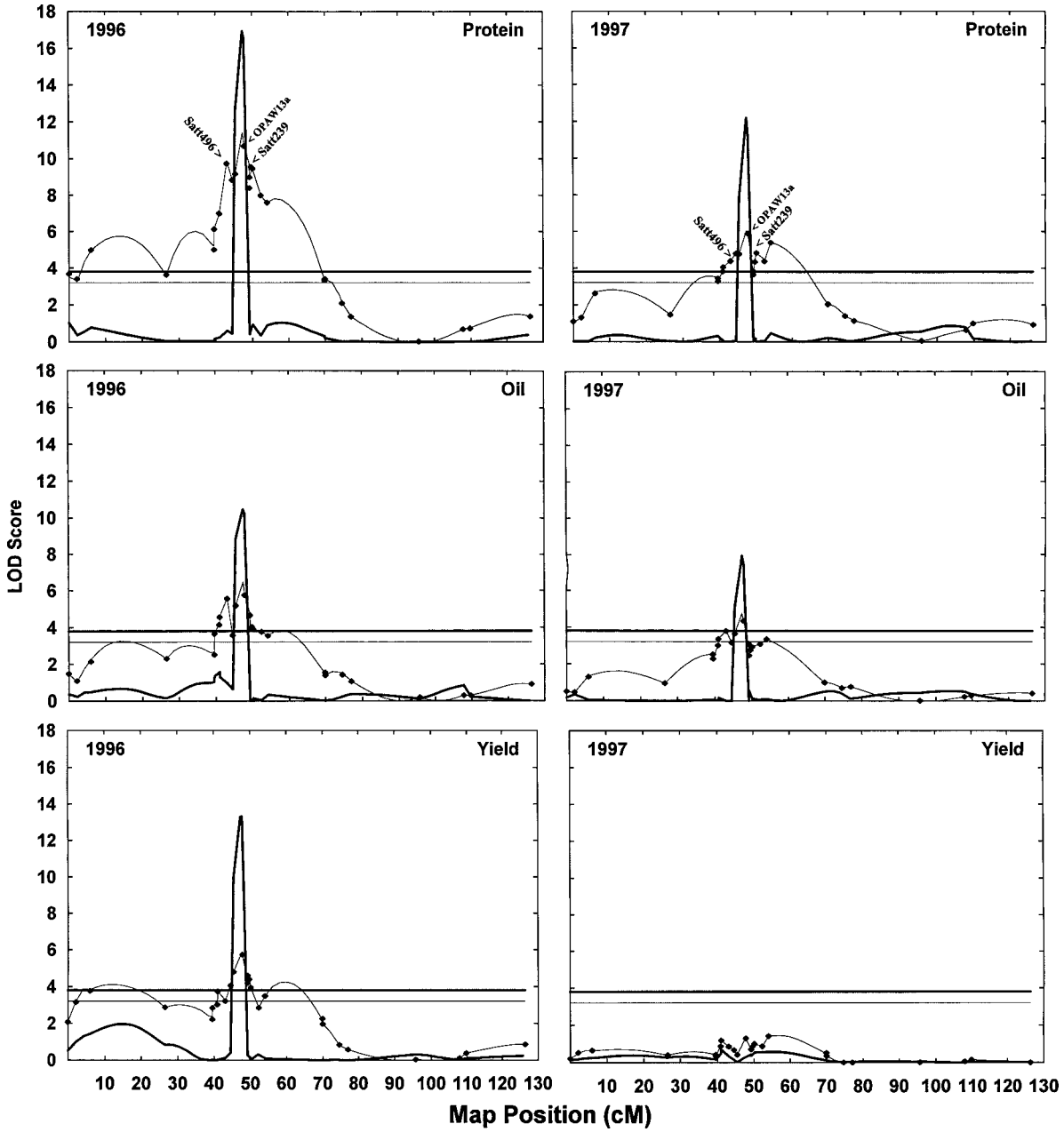



**Fig. 3. A graph of LOD score versus map distance (in centimorgans) in the upper portion of soybean linkage group I for seed protein, oil, and yield in 1996 and 1997. The LOD scores were derived from linear regression (LR- diamond symbols), simple interval mapping (SIM-thin lines), and composite interval mapping (CIM-thick lines) analyses. The permutation-derived ($n$ = 1000 per trait) LOD score significance criteria were 3.4, 3.2 (horizontal thin lines), and 3.8 (horizontal thick lines), respectively. For the protein QTL plot, the effects of the closest dominant RAPD marker OPAW13a and closest flanking SSR markers (Satt496 and Satt239) are indicated.**

In contrast to SIM-identified major and minor peaks on LG-I, the CIM-based scan revealed only one QTL of strong effect (Fig. 3). In effect, when the LG-I marker associated with the primary QTL was used as a BG marker (to control trait variation), the other LG-I markers had no independent effect on the trait. Evidently, the SIM-identified minor QTLs flanking the major QTL were statistical artifacts.

The number of markers identified by stepwise regression as having a $P = 0.01$ effect on the seed protein and oil variation in 1996 and 1997 ranged from seven to eight (Table 4), including of course the LG-I marker most highly associated with the protein and oil QTL. Eleven markers were identified in the 1996 yield data, but only five markers (none on LG-I) were identified in the 1997 yield data. About 65 to 85% of the total variation in protein and oil each year, and yield and maturity in 1996, could be accounted for with a statistical model of just one LG-I QTL conditioned on the variance governed by the indicated number of background markers ($TR^2$ data in Table 4). About 24 to 28% of the trait variation was controlled by each LG-I QTL itself ($R^2$ data in Table 4), excepting 1996 protein (45%), 1996 yield (19%), 1996 maturity (10%), and 1997 yield (0%).

**Table 4. Statistical parameters derived for the protein, oil, and yield QTLs identified on soybean linkage group I using the composite interval mapping (CIM) module of QTL Cartographer (Basten et al., 2001). Map positions reflect centimorgan distance from terminal marker Satt451 (Fig. 1). The bracketing markers are those markers immediately external of the left and right flanks of the confidence interval surrounding the mean map position of the QTL.**

| Year and trait | Left bracketing marker name (cM) | Left flank of QTL (cM) | Mean QTL map position (cM) | Right flank of QTL (cM) | Right bracketing marker name (cM) | QTL LOD score peak | Allele A additive effect † | Number of back-ground markers‡ | $R^2$ ‡ (%) | $TR^2$ ‡ (%) | Joint map test statistic | QTL×Y test statistic | Mean (2-yr) additive effect† |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | 1996–1997 Protein | |
| 1996 Protein | OPAJ13a 45.45 | 45.48 | 47.45 | 47.81 | OPAW13a 47.82 | 16.92* | −1.00 | 7 | 45.3 | 76.8 | | | |
| 1997 Protein | OPAJ13a 45.45 | 45.48 | 47.45 | 47.81 | OPAW13a 47.82 | 12.18* | −0.89 | 8 | 27.7 | 77.5 | 20.48* | 1.31* | −0.92 |
| | | | | | | | | | | | | 1996–1997 Oil | |
| 1996 Oil | OPAJ13a 45.45 | 45.48 | 47.45 | 47.81 | OPAW13a 47.82 | 10.46* | +0.57 | 8 | 27.8 | 71.1 | | | |
| 1997 Oil | OPAJ13a 45.45 | 45.48 | 47.45 | 47.81 | OPAW13a 47.82 | 7.94* | +0.59 | 7 | 24.2 | 65.6 | 13.95* | 0.00 ns | +0.57 |
| | | | | | | | | | | | | 1996–1997 Yield | |
| 1996 Yield | OPAJ13a 45.45 | 47.81 | 47.82 | 49.23 | OP_L04b 49.26 | 13.31* | +134 | 11 | 19.4 | 84.4 | | | |
| 1997 Yield | – | – | – | – | – | 0.0 | – | 5 | 0.0 | 0.0 | 12.16* | 2.02* | +111 |
| 1996 Maturity | OP_R10 44.55 | 44.55 | 44.55 | 45.43 | OPAJ13a 45.45 | 5.46* | +1.55 | 8 | 10.2 | 74.2 | | NA (no 1997 data) | |

* Significant LOD score or test statistic at a genomewise probability level of $\alpha = 0.05$. See Materials and Methods for computational details.
† The units for the allele A additive effect are 10g $kg^{-1}$ (i.e., percentage of protein and oil in the seed), kg $ha^{-1}$ for yield, and d for maturity.
‡ The number of markers identified for each trait, via a stepwise forward then backward regression analysis, as having a significant effect ($P = 0.01$) on the trait variance. $R^2$ is the proportion of total trait variance explained by the QTL itself (at its map position) on control of the variation governed by the background markers. $TR^2$ is the proportion of total trait variance explained by both the QTL and by the background (BG) markers (none of these were significant at the QTL level).

CIM is essentially a method for enhancing the signal-to-noise ratio, where the signal is an allelic effect of a QTL judged to be significant at an appropriate LOD threshold. A CIM analysis thus treats as noise all trait variation arising from significant QTLs in other parts of the genome and/or from genetic markers whose effect on the trait, while measurable, is not statistically sufficient in terms of exceeding the given LOD threshold for a QTL declaration.

The QTLs detected in the 1996 and 1997 protein and oil data mapped to a coincident position (47.45 cM) on LG-I (Table 4). These QTLs were bracketed by the same two dominant RAPD markers, OPAJ13a (45.45 cM) and OPAW13a (47.82 cM), and were bounded by an identical confidence interval of 2.33 cM (i.e., 45.48–47.81 cM). The 1996 yield QTL mapped to the LG-I position (47.82 cM) occupied by marker OPAW13a, thus making RAPD marker OP_L04b (49.26 cM) its rightmost bracketing marker (Table 4). This yield QTL was bounded by a somewhat smaller confidence interval of 1.42 cM (47.81–49.23 cM). The CIM analysis detected no QTLs in the 1997 yield data. A significant LG-I QTL for 1996 maturity was identified in the CIM (but not the SIM) analysis. This maturity QTL mapped to a position coincident with that of RAPD marker OP_R10 (44.55 cM), which along with OPAJ13a (45.45 cM), bracketed this QTL. It also had a rather narrow confidence interval of 0.88 cM (44.55–45.43 cM). Note that the LG-I map position of the 1996 yield QTL (47.82 cM) fell just barely outside of the 1996 protein and oil QTL confidence interval (45.48–47.81 cM), which in turn was very close to the confidence interval (44.55–45.43 cM) of the 1996 maturity QTL.

The CIM-estimated additive effects of allele A (i.e., the Asgrow A3733 allele) at the LG-I protein and oil QTL(s) were a respective –1.00 and +0.57 (1996) and –0.89 and +0.59 (1997) percentage points (Table 4). In comparison, Diers et al. (1992) reported that the *G. max* segment of LG-I segregating in their *G. max* × *G. soja* population had protein/oil additive effects of about −1.10/+0.75, respectively, whereas Sebolt et al. (2000) reported values of −0.975/+0.45 for the same segment. In the present study, the ratios of the protein/oil additive effects for the LG-I protein and oil QTL(s) were −1.70 (1996) and −1.52 (1997), which were similar to −1.77 and −1.51 values computed for the 1996 and 1997 ratios of the protein/oil phenotypic regression coefficients (Fig. 2). The additive effect of allele A at the LG-I yield QTL was a statistically significant +134 kg $ha^{-1}$ in 1996 (Table 4), but was not detectable in 1997 becaue of the yield measurement imprecision that year.

The joint mapping module of QTL Cartographer was applied to the 1996 and 1997 data for protein, oil, and yield to assess the significance of any CIM-based QTL × Y interactions. The QTL × Y interaction for protein was nominally significant, arising from a negative additive effect of QTL allele A on seed protein content that was somewhat greater in 1996 than in 1997. The significant QTL × Y interaction for seed yield was the result of a large positive effect of QTL allele A on yield in 1996, but an essentially zero effect in 1997.

## DISCUSSION

The LG-I QTLs for protein, oil, and yield mapped to an interval flanked by the codominant SSR markers Satt496 and Satt239 (Fig. 1). The two-point map distance between these SSRs was 0.85 cM in this population of 76 RILs. These two SSRs were completely linked (0.0 cM) in the LG-I map of the 59-plant $F_2$ mapping population of the *G. max* × *G. soja* mating reported in Cregan et al. (1999). Five dominant RAPD markers mapped to positions within the Satt496- Satt239 interval. The dominant (+) alleles of RAPD markers OPAJ13a, OPAW13a, and OPAX03 originated from the Asgrow A3733 parent, whereas the dominant (+) alleles of OP_R10 and OP_L04b originated from the PI 437088A parent. Maximum likelihood estimates of linkage between repulsion-phased dominant markers are known to be biased upwards (Liu, 1998), and in any $F_5$–derived RIL population, residual heterozygotes at dominant marker loci are unavoidably misclassified with their dominant homozygotes. As a result, the five RAPD markers inflated the Satt496– Satt239 interval map interval about 8-fold, to 6.63 cM. Still, the protein, oil, and yield QTLs mapped very near OPAW13a, and its two-point map distance with Satt239 was 0.84 cM. The LG-I RFLPs K011 and A407, identified in the marker association analyses conducted by Diers et al. (1992) as having the strongest and second strongest associations with protein and oil, map 1.8 and 1.7 cM above Satt239 (Cregan et al., 1999). Brummer et al. (1997) reported that RFLP markers A407 and A144 had strong marker associations with protein, and Sebolt et al. (2000) presented a SIM-based LOD score scan that showed QTLs for protein and oil peaking near A144, which maps 7.1 cM above Satt239 (Cregan et al., 1999). Satt700, an SSR tightly linked with Satt496, was created from the DNA sequence of a clone pulled from a BAC library on the basis of sequence homology with RFLP clone L185 (Shoemaker and Cregan, unpublished data). The RFLP marker L185 resides 3.6 cM above Satt239 (Cregan et al., 1999).

QTL interval analyses indicated that seed protein was increased by 1.84 percentage points on the substitution of two PI 437088A alleles (i.e., BB) for two Asgrow A3733 alleles (i.e., AA) at OPAW13a, the marker nearest the LG-I protein QTL (Table 4). However, that same substitution coordinately depressed seed oil by 1.14 percentage points, and in one of the trial years, depressed seed yield by 268 kg $ha^{-1}$ (3.9 bu $ac^{-1}$). These effects are depicted in Fig. 2, where solid and open symbols identify RILs with respective AA or BB genotypes at the locus OPAW13a. In 1996 trial, RILs of the BB genotype (i.e., open symbols) fall mainly in the left side of the graph, consistent with their lower yield, but greater protein and lower oil, compared to RILs of the AA genotype (i.e., solid symbols). Only the protein and oil pattern was repeated in the 1997 trial, presumably because of imprecision in the yield measurement that year.

No RIL with a recombinant coupling-phased phenotype of high protein – high oil (or its alternative of low protein – low oil) was clearly identifiable in Fig. 2. This outcome was consistent with the null hypothesis of a single QTL conditioning (inverse) pleiotropic effects on protein and oil. The alternative hypothesis, that a protein QTL is in tight (repulsion phase) linkage with an oil QTL, is less parsimonious. Its acceptance requires one to pose an additional hypothesis of what the true recombination fraction between two such QTLs might be (Hanson, 1959), given that no (i.e., $x = 0$) recombinants were actually observed in the RIL population (i.e., $n = 76$). One might reasonably assume the recombination fraction ($R = x/n$) is less than one recombinant per 76 RILs. Because of additional selfing, the recombination fraction ($R$) in an RIL population will be greater than that ($r$) of an $F_2$ population. Using the equation $r = R/(2\text{-}2R)$ derived by Haldane and Waddington (1931) for selfed (not sib-mated) RIL populations, an RIL recombination fraction ($R$) of $< 1/76$ translates into an $F_2$ recombination fraction ($r$) of $<0.0067$. When transformed into Haldane map units, i.e., $m = -0.5 \times ln\ (1 - 2r)$, the linkage distance between two LG-I QTLs, one for protein and one for oil would clearly have to be less than 0.67 cM.

A QTL with inversely pleiotropic effects on protein and oil is consistent with the results of Wilcox and Cavins (1995) and Wilcox (1998). In those studies, the linear coefficients for protein regressed on oil strengthened from −1.51 to −1.72 during backcross introgression of the high protein allele from Pando into the high yield recurrent parent, and from −1.55 to −1.75 during cycles five to eight of recurrent selection for high protein. The QTL allelism test conducted by Sebolt et al. (2000) demonstrated that the *G. soja* (PI 468916) allele for high protein and low oil was allelic with the Pando allele for high protein and low oil. Thus, it is now clear that in the Pando populations, the genotypic-level regression coefficients were simply reflective of the ratio of the additive effects exerted by the Pando (= *G. soja*) allele on soybean seed protein and oil. We have not yet confirmed the allelism of the Pando allele with the PI 437088A allele. However, both alleles map to the same small segment of LG-I, and the latter allele has protein/oil additive effect ratio of about –1.51 to −1.77, which is certainly consistent with a hypothesis of allelism.

The 1996 CIM analysis indicated that a QTL for maturity might reside near the QTL(s) for protein, oil, and yield. The map position and strength of this maturity QTL will need to be verified, since only two of the six subplots per replicate were scored for maturity in 1996, and 1997 maturity data were not reliable (i.e., green stems). A physiological coupling of earlier senescence with higher soybean seed protein would not be entirely unexpected, given the "self-destruct" hypothesis of Sinclair and de Wit (1975). These authors hypothesized that soybean, because of its high seed protein content, is unable to meet the massive demand for N (during seed filling) from just soil N uptake and $N_2$ fixation, and must thus remobilize N from its vegetative tissues. An early "shutdown" of carbon assimilation would be expected to hasten senescence.

We conclude with a discussion of the allelic effects

of the LG-I QTL and the correlation of yield with seed protein (negative) and seed oil (positive) observed in our study (and in many other studies). In a now classic paper, Hanson et al. (1961) examined the genetic basis of the inversely coordinate genetic and environmental variation in soybean seed protein and oil. The authors noted that the average calorific values for soybean oil and protein were about 9.4 and 4.6 Kcal $g^{-1}$, which translated into an oil/protein energy ratio of about 2.0 (i.e., 9.4/4.6). The synthesis of oil or protein requires energy, which can be obtained by oxidizing (some of the) carbons present in carbohydrate. However, the authors noted that seed carbohydrate changed little on genetic (or environmental) alteration of the seed protein (or oil) content. Indeed, increases in protein were accompanied by coordinate decreases in oil, leading the authors to speculate that the synthetic pathways for protein and oil compete for the same carbon and energy sources. Hanson et al. (1961) computed that, on a nongenetic (environmental) scale, the protein/oil exchange ratio was nearly equivalent to the intrinsic calorific ratio of 2.0 (i.e., 1.92 unit of seed protein was gained per 1.0 unit loss in seed oil). However, on a genotypic scale, computed from a large set of soybean genotypes, the calculated protein/oil exchange ratio was substantially less, about −1.5 or −1.6. This finding led the authors to conclude that if breeding lines high in seed oil were to be genetically converted into lines high in seed protein, at a protein/oil exchange ratio less than a calorific ratio of two, then the high protein lines should be higher yielding. Although Hanson et al. (1961) were aware of prior reports of negative correlations of protein with yield, and in fact observed a negative correlation in their own data, they still concluded that high protein and high yield were compatible. Since their report, however, negative genotypic correlations between yield and protein have been repeatedly documented in the soybean literature. We now show that the PI 437088A allele of a LG-I QTL positively affects protein, but negatively affects oil and yield. The energetic cost associated with increased protein deposition (at the expense of oil) in the soybean seed via genetic (but apparently not environmental) means would appear to be seriously underestimated. Cloning this QTL would prove useful, if only to learn how it functionally mediates the allocation of photosynthetic carbon and energy between seed protein, oil, and yield.

## REFERENCES


Akkaya, M.A., A. Bhagwat, and P.B. Cregan. 1992. Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132: 1131–1139.

Allard, R.W. 1956. Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24:235–278.

Basten, C.J., B.S. Weir, and Z.-B. Zeng. 2001. QTL Cartographer. Version 1.15. A reference manual and tutorial for QTL mapping. Department of Statistics. North Carolina State University, Raleigh, NC.

Brim, C.A. 1973. Quantitative genetics and breeding. p. 155–186. *In* B.E. Caldwell (ed.) Soybeans: Improvement, production, and uses. ASA, Madison, WI.

Brim, C.A., and J.W. Burton. 1979. Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. Crop Sci. 19: 494–498.

Brummer, E.C., G.L. Graef, J. Orf, J.R. Wilcox, and R.C. Shoemaker. 1997. Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37:370–378.

Burton, J.W. 1987. Quantitative genetics: Results relevant to soybean breeding. p. 211–247. *In* J.R. Wilcox (ed.) Soybeans: Improvement, production, and uses. 2nd Edition. ASA, Madison, WI.

Churchill, R.W., and G.A. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963–971.

Cober, E.R., and H.D. Voldeng. 2000. Developing high-protein, high-yield soybean populations and lines. Crop Sci. 40:39–42.

Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, T.T. Van Toai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39:1464–1490.

Csanádi, G., J. Vollmann, G. Stift, and T. Lelly. 2001. Seed quality QTLs identified in a molecular map of early maturing soybean. Theor. Appl. Genet. 103:912–919.

Diers, B.W., P. Keim, W.R. Fehr, and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83:608–612.

Ferreira, A.R., K.R. Foutz, and P. Keim. 2000. Soybean genetic map of RAPD markers assigned to an existing scaffold RFLP map. J. Hered. 91:392–396.

Haldane, J.B.S., and C.H. Waddington. 1931. Inbreeding and linkage. Genetics 16:357–374.

Hanson, W.D. 1959. Minimum family sizes for the planning of genetic experiments. Agron. J. 51:711–715.

Hanson, W.D., R.C. Leffel, and R.W. Howell. 1961. Genetic analysis of energy production in the soybean. Crop Sci. 1:121–126.

Hartwig, E.E. 1973. Varietal development. p. 187–210. *In* B.E. Caldwell (ed.) Soybeans: Improvement, production, and uses. ASA, Madison, WI.

Hartwig, E.E., and K. Hinson. 1972. Association between chemical composition of seed and seed yield of soybeans. Crop Sci. 12: 829–830.

Hartwig, E.E., and T.C. Kilen. 1991. Yield and composition of soybean seed from parents with different protein, similar yield. Crop Sci. 31:290–292.

Helms, T.C., and J.H. Orf. 1998. Protein, oil, and yield in soybean lines selected for increased protein. Crop Sci. 38:707–711.

Herman, E.M. and B.A. Larkins. 1999. Protein storage bodies and vacuoles. Plant Cell 11:601:614.

Howell, R.W., and J.L. Cartter. 1958. Physiological factors affecting composition of soybeans. II. Response of oil and other constituents of soybeans to temperature under controlled conditions. Agron. J. 50:664–667.

Hurburgh, C.R. 2001. Quality of the 2001 soybean crop from the United States. American Soybean Association. http://www.exnet.iastate.edu/Pages/grain/test/soybean/01sbqual.pdf (Verified 3 December 2002).

Jiang, C., and Z.-B. Zeng. 1995. Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140:1111–1127.

Johnson, H.W., H.F. Robinson, and R.E. Comstock. 1955. Genotypic and phenotypic correlations in soybeans and their implications in selection. Agron. J. 47:477–483.

Keim, P., J.M. Schupp, S.E. Travis, K. Clayton, T. Zhu, L. Shi, A. Ferreira, and D.W. Webb. 1997. A high-density soybean genetic map based on AFLP markers. Crop Sci. 37:537–543.

Knapp, S.J., W.W. Stroup, and W.M. Ross. 1985. Exact confidence intervals for heritability on a progeny mean basis. Crop Sci. 25: 192–194.

Lander, E.S., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199.

Lee, S.H., M.A. Bailey, M.A.R. Mian, T.E. Carter, Jr., E.R. Shipe, D.A. Ashley, W.A. Parrott, R.S. Hussey, and H.R. Boerma. 1996. RFLP loci associated with soybean seed protein and oil content across populations and locations. Theor. Appl. Genet. 93:649–657.

Lincoln, S.E., and S.L. Lander. 1993. Mapmaker/exp 3.0 and Mapmaker/QTL 1.1. Whitehead Inst. of Med. Res. Tech. Report. Cambridge, MA.

Liu, B.H. 1998. Statistical genomics: linkage, mapping, and QTL analysis. CRC Press LLC, Boca Raton, FL.

Mansur, L.M., K.G. Lark, H. Kross, and A. Oliveira. 1993. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (Glycine max L.). Theor. Appl. Genet. 86:907–913.

Mansur, L.M., J.H. Orf, K. Chase, T. Jarvik, P.B. Cregan, and K.G. Lark. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci. 36:1327–1336.

Mode, C.J., and H.F. Robinson. 1959. Pleiotropism and the genetic variance and covariance. Biometrics 15:518–537.

NGRP. 2001. National Genetic Resources Program, USDA-ARS. Germplasm Resources Information Network (GRIN). Online database of the NGR Laboratory, Beltsville, Md. http://www.ars-grin.gov/npgs/descriptors/soybean (Verified 3 December 2002). http://www.ars-grin.gov/cgibin/npgs/html/acc_search.pl?accid = PI+437088A

NCBI. 2001. National Center for Biotechnology Information. GenBank. http://www3.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db = 2&form = 1&term = D28876 (Verified 1 June 2002).

Olsen, H.L. 2001. Precision mapping of a QTL affecting soybean seed protein, oil, and yield. M.S. thesis. Univ. of Nebraska, Lincoln.

Orf, J.H., K. Chase, T. Jarvik, L.M. Mansur, P.B. Cregan, F.R. Adler, and K.G. Lark. 1999. Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. Crop Sci. 39:1642–1651.

Panford, J.A. 1987. Application of near-infrared reflectance spectroscopy in North America. p. 201–211. *In* P. Williams and K. Norris (ed.) Near-infrared technology in the agricultural and food sciences. Assoc. Cereal Chem. Inc., St. Paul, MN.

Qiu, B.X., P.R. Arelli, and D.A. Sleper. 1999. RFLP markers associated with soybean cyst nematode resistance and seed composition in a 'Peking' × 'Essex' population. Theor. Appl. Genet. 98:356–364.

Rongwen, J., M.S. Akkaya, A.A. Bhagwat, U. Lavi, and P.B. Cregan. 1995. The use of microsatellite DNA markers for soybean genotype identification. Theor. Appl. Genet. 90:43–48.

Saghai-Maroof, M.A., K.M. Soliman, R.A. Jorgensen, and R.W. Allard. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. (USA) 81:8014–8018.

SAS. 1999. What's New in SAS® Software for Version 8. SAS Institute Inc., SAS Campus Drive, Cary, NC; http://www.sas.com/service/library/onlinedoc/v8/whatsnew (Verified 3 December 2002).

Schwenk, F.W. and C.D. Nickell. 1980. Soybean green stem caused by bean pod mottle virus. Plant Dis. Rep. 64:863–865.

Sebern, N.A., and J.W. Lambert. 1984. Effect of stratification for percent protein in two soybean populations. Crop Sci. 24:225–228.

Sebolt, A.M., R.C. Shoemaker, and B.W. Diers. 2000. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Sci. 40:1438–1444.

Shannon, J.G., J.R. Wilcox, and A.H. Probst. 1972. Estimated gains from selection for protein and yield in the $F_4$ generation of six soybean populations. Crop Sci. 12:824–826.

Shoemaker, R.C., and J.E. Specht. 1995. Integration of the soybean molecular and classical genetic linkage maps. Crop Sci. 35:436–446.

Sinclair, T.R., and C.T. de Wit. 1975. Photosynthate and nitrogen requirements for seed production by various crops. Science 189: 565–567.

Skroch, P., and J. Neinhuis. 1995. Impact of scoring error and reproducibility of RAPD data on RAPD based estimates of genetic distance. Theor. Appl. Genet. 91:1086–1091.

SoyBase. 2002. SoyBase– a genome database for Glycine. Administered online by USDA-ARS and Iowa State University. http://129.186.26.94/SSR.html (Verified 3 December 2002).

Specht, J.E., K. Chase, M. Macrander, G.L. Graef, J. Chung, J.P. Markwell, M. Germann, H.H. Orf, and K.G. Lark. 2001. Soybean response to water: A QTL analysis of drought tolerance. Crop Sci. 41:493–509.

Specht, J.E., J.H. Williams, and C.J. Weidenbenner. 1986. Differential responses of soybean genotypes subjected to a seasonal soil water gradient. Crop Sci. 26:922–934.

Thorne, J.C., and W.R. Fehr. 1970. Incorporation of high-protein, exotic germplasm into soybean populations by 2- and 3-way crosses. Crop Sci. 10:652–655.

Thompson, S., S. Sonka, E. Nafziger, M. Westgate, P. Khanna, P. Orwick, R. Esgar, and D. Sigberg. 2001. Varietal information program for soybeans. An online analytical tool. http://web.aces.uiuc.edu/vips/v2home/vips2home.cfm (Verified 3 December 2002).

Wehrmann, V.K., W.R. Fehr, S.R. Cianzio, and J.F. Cavins. 1987. Transfer of high seed protein to high-yielding soybean cultivars. Crop Sci. 27:927–931.

Wilcox, J.R. 1998. Increasing seed protein in soybean with eight cycles of recurrent selection. Crop Sci. 38:1536–1540.

Wilcox, J.R., and J.F. Cavins. 1995. Backcrossing high seed protein to a soybean cultivar. Crop Sci. 35:1036–1041.

Wilcox, J.R., and Z. Guodong. 1997. Relationships between seed yield and seed protein in determinate and indeterminate soybean populations. Crop Sci. 37:361–364.

Williams, J.G.K., A.R. Kubelic, K.J. Livack, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531–6535.

Yamanaka, N., S. Ninomiya, M. Hoshi, Y. Tsubokura, M. Yano, Y. Nagamura, T. Sasaki, and K. Harada. 2001. An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology, and regions of segregation distortion. DNA Research 8:61–72. (an online journal) http://dna-res.kazusa.or.jp/8/2/02/HTMLA/ (Verified 3 December 2002).

Zeng, Z. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457–1468.